

AI安全非小事 風險管理須完善

美東時間9月3日，美國三大人工智能平台——OpenAI的ChatGPT、Anthropic的Claude以及SpaceXAI的Grok罕見地在同一時間內發生集體「宕機」事件，導致服務中斷4小時。這場被形容為AI「黑色星期四」的同步癱瘓事件，為全球敲響警鐘。加強人工智能領域的治理，完善相關立法及執行機制，確保其在符合安全前提下的有序運作，是擺在當前的最迫切任務。

三家機構事後公布的原因各有不同，包括路由器錯誤、基礎設施問題、數據中心故障等，但共同指向基礎設施不穩定、不可靠的行業弊端。目前，人工智能平台層出不窮，彼此是競爭對手，但這些平台高度依賴極少數雲端基建與網絡安全供應商，當大量使用者集中訪問、導致底層公有雲基礎設施出現微小波動時，看似獨立的競爭對手會瞬間同步「斷氣」。與此形成鮮明對比的是，擁有自家雲端與網絡架構的一些人工智能平台，在事件中受影響最小，這反向證明了基礎設施高度集中的巨大隱患。

當幾大AI平台同步癱瘓時，在社交媒體上瞬間引發了全球用戶的焦慮和混亂，這揭示了企業和個人對AI的依賴與日俱增，一旦AI出故障，很多人不知所

措。以香港為例，根據科大早前的調查，高逾7成2在職人士每周或每日使用AI工具，甚至形成職業依賴，但接受過系統AI應變培訓的人少之又少，一旦類似事件重演，許多人在沒有AI輔助下失去獨立工作的能力。

人工智能會不會取代人，或者讓大多數人變成人工智能的奴隸，早已引起各方憂慮。這是倫理問題，也是現實的嚴峻挑戰。應高度重視人工智能引發的各種內生和衍生風險，確保人工智能始終在人類控制之下，成為人類可信的工具，這需要加強基礎設施建設，加快相關立法，提升人們在突發事故下的應變能力。

今次事故還引發一個更深層次的問題。在國際地緣政治日益複雜化的情況下，人工智能「武器化」傾向日益明顯。美國泛化國家安全概念，從算力晶片、雲端訪問、資金投資、終端產品乃至模型審查等多個核心維度，全方位打壓中國人工智能發展。在中國推動科技獨立自主、「DeepSeek時刻」不斷出現的情況下，極端反華勢力非常焦慮。華盛頓智庫「新美國安全中心」曾發表報告，建議美國國防部與國安機構預先研擬情報與作戰方案，甚至提議在中國

通用人工智能取得突破前夕，美方應考慮對中國運行AI的數據中心實施動能打擊（軍事轟炸）或網絡攻擊，以阻止中國技術領先。難怪有人指出，美國三大智能平台同時癱瘓，相當於一次有目的的「軍事演習」。

人工智能安全關係國家安全，這是一個客觀而現實的問題。中國作為負責任大國，在人工智能領域始終致力於做國際公共產品的提供者，源源不斷貢獻中國方案，包括為發展中國家培訓技術專家。而在人工智能治理方面，「一國兩制」下的香港可以發揮獨特作用。一方面，香港可以對接內地算力資源和應用場景，在AI應用上發揮協助「引進來、走出去」的雙向作用；另一方面，憑藉香港的國際化優勢和普通法制度，建構安全可控的AI治理機制，加強國際標準的對接，協助內地開源生態的全球嵌入深度。

科技是把雙刃劍，既能載舟亦能覆舟。香港首個五年規劃即將發布，要抓住AI時代帶來的巨大機遇，加快推動產業AI化、AI產業化，同時也要加強風險管理，在構建AI治理體系方面先行先試。這不僅能造福香港，也為提升國家在全球AI治理中的主導作用作貢獻。

井水集

保障市民健康

為了迎合消費者「既想食甜又怕長胖」的心理，市面上出現越來越多聲稱「無糖」的食品。實際上，這些食品大多是利用甜味劑（代糖）來提供甜味，而代糖種類繁多，過量食用一樣會引發健康問題。特區政府昨日就《食物內甜味劑規例》修訂建議展開公眾諮詢，接軌最新國際標準，這是保護公眾健康的應有之舉。

由聯合國糧農組織和世衛組織成立的食物法典委員會不時更新《食物添加劑通用標準》，這也是香港今次修訂的主要參考標準。今次修訂有三項重點內容：一是建議將10多種「多羥醇」納入規管。「多羥醇」常見於香口膠、朱古力、氣泡水、棒棒糖、止咳糖漿、夾心餅乾、蜜餞、果凍等食品中。港大研究曾指出，很多意想不到的食品包括鹹食中都含有甜味劑，消費者或在不知不覺中過量攝入，並造成腹瀉、胃氣脹、腹脹等問題。政府建議將之納入監管，是與時俱進的體現。

二是為甜味劑訂立含量上限。目前的監管制度下，只有獲

准的甜味劑名單，但沒有限定含量，在食品生產過程中容易過量使用。作出技術性修訂後，甜味劑含量與食材分量相對應，消費者健康更有保障。

三是擴大准用甜味劑名單，讓業界與消費者有更多選擇。代糖的生產和工藝在不斷改進，如新一代甜味劑「愛德萬甜」，其甜度是蔗糖的2萬倍，這意味着只要用很少含量的代糖，就能達到「甜」的效果，消費者因此減少了代糖的攝入。由於新一代甜味劑被廣泛使用，香港需要擴大名單，才能讓更多優質的新產品進入市場。

香港食品監管一向以嚴格著稱，但在添加劑不斷更新迭代的情況下，香港有必要緊跟國際標準，不斷優化監管制度。甜味劑只是添加劑的一種，其他如防腐劑、調味劑等一樣要提升監管力度，加強抽查、檢測能力。在香港食品九成以上進口的情況下，加強與大灣區等主要食品來源地的聯合監管，正是大灣區融合的應有之義。

國際監管聯手 提升金融AI安全

BIS與香港英國沙特研工具包 評估AI用例可靠性

金融科技

人工智能

（AI）在金融機構包括銀行業的應用不斷擴大和深化，與AI相關的風險管理也值得重視。香港金融管理局總裁余偉文表示，除了透過生成式人工智能沙盒（GenAI Sandbox）促進金融服務領域的創新，金管局有份參與的AI工具包（Project Noor）有望於下一季度開始測試，用以評估個別AI用例（Use Case）是否可靠，而且更重要的是，是否具備可解釋性（Explainable）。

大公報記者 黃裕慶

Project Noor由國際結算銀行（BIS）創新樞紐轄下的香港中心、香港金融管理局、英國金融行為監管局（FCA），以及沙特阿拉伯中央銀行合作開發，旨在於受控環境中對最新的可解釋AI（Explainable AI）技術展開原型驗證。

據BIS網頁資料所載，Project Noor旨在為金融監管機構提供獨立、實用的工具，以協助評估並解讀銀行及其他金融機構所使用的AI模型的內部運作機制。

了解AI判斷背後邏輯

舉例說，AI模型現時已能協助核准按揭貸款、設定信用卡額度，並即時偵測潛在欺詐。儘管這些服務看起來順暢無縫，但要理解模型為何判定「可以」、「不可以」或「需進一步審查」時，仍可能讓人覺得難以看清。

根據新法規要求，BIS指高風險的金融AI必須具備可解釋性與可審計性（Explainable and Auditable），但監管機構目前仍缺乏一套共同且實用的操作手冊，而於2025年8月提出的Project Noor，可望提供一個原型（Prototype）或AI工具，使監管機構能夠驗證模型的透明度、評估其公平性，並測試其穩健性。

維持數碼金融公平性

余偉文昨日在香港銀行公會年度卓越人士講座暨午餐會上表示，上述AI工具包在經過測試之後，期望可於明年初推出，屆時將可提供予銀行業界使用。BIS指出，AI模型若具備清楚、可供人閱讀理解的人類化解釋，能夠增強信任，並幫助維持數碼金融

的公平性。

香港金管局近年積極推動業界向智慧型銀行的方向轉型，為鼓勵銀行業更廣泛採用AI，同時在創新、風險管理與市場信心之間取得適當平衡，余偉文指金管局的另一個工作方向，是透過生成式人工智能沙盒（並已於今年3月擴大為GenAI沙盒++），在風險可控的環境中開發、測試及優化其用例，推動香港金融生態系統負責任地採用AI。

GenAI沙盒++繼續聚焦於三大重要範疇，即風險管理、反詐騙及客戶體驗，此外也會持續推進「AI對抗AI」策略，運用AI來管理AI應用所帶來的風險。余偉文表示，GenAI沙盒++涵蓋多個金融領域，包括銀行、證券及資本市場、資產與財富管理、保險、強積金及儲值支付工具，此外也會透過平台與香港、內地的科技公司（尤其粵港澳大灣區）、解決方案供應商等分享發展經驗，藉此共建健康的生態圈。

最後，余偉文表示，金管局也攜同香港銀行公會、香港銀行學會等機構，致力為從業員提供與AI相關的培訓。他指出，在應用AI過程中，相信仍要以某種方式保留人的要素，以作適當監察。



▲余偉文表示，期望銀行業界可於明年初使用Project Noor。
大公報記者林少權攝

▲香港金管局與BIS、英國FCA、沙地中央銀行合作開發的Project Noor，將於下一季度開始測試。
資料圖片



提升應用AI信心
金管局三方面

1 透過使用「生成式人工智能」（GenAI）沙盒，便利業界在風險可控、環境安全的情況下，測試各種AI工具和用例

2 與BIS、英國FCA和沙地中央銀行合作，透過「Project Noor」研發工具，以協助監管機構評估並解讀銀行及其他金融機構所使用的AI智慧模型的內部運作機制

3 在使用AI過程中保留「人的要素」（human element），以某種形式來監督AI的使用

大公報AI製圖

證監擬向券商發AI應用指引

【大公報訊】證監會梁鳳儀昨日在香港銀行公會年度卓越人士講座暨午餐會的專題討論環節中表示，香港作為國際金融中心，人工智能（AI）帶來三大主要裨益：提升投資者成果、深化資本形成，以及賦能中介機構創造更大價值。透過AI，投資者可運用更有效的工具加深理解資訊、獲取優質投資建議、比較不同選項，並作出更適當的決策。此外，香港正迎來龐大機遇，將AI價值鏈的不同企業與國際、內地及區域資本對接。投資者亦可借助香港，分享戰略性科技產業增長所帶來的紅利。

隨著AI在證監會持牌機構中的應用場景日益增加，在妥善管控下，AI能提升中介機構的效率、一致性及客戶服務水平，同時讓人力資源轉向更高價值的策略判斷及持份者溝通工作。梁鳳儀表示，證監會的監管方針

聚焦於安全及負責任地使用AI，並基於現行監管原則，這包括高級管理層問責、妥善管治、充足的系統和監控、高透明度，以及保障客戶利益。證監會正考慮是否需要提供進一步指引，協助機構在代理式AI的應用場景中，落實現有監管原則。

須明確劃分責任 實行有效監督

指引並非為規範技術設計細節，而是為明確劃分責任及實行有效的人類監督、設立受控的存取權限及訂明應用範圍、在部署前進行穩健測試並持續監察、確保營運韌性，以及在出現問題時迅速上報和採取補救措施。證監會持續進行數字轉型，以提升監管效率、市場監察能力及營運韌性，相關效益包括縮短牌照處理時間，以及更早偵測失當行為和詐騙活動。

事實上，證監會對AI在金融業的

應用採取「雙軌並行」的監管方針，既支持技術創新，亦嚴格防範潛在風險。例如在2024年底響應特區政府發表的《有關在金融市場負責任地應用人工智能的政策宣言》，發布針對生成式AI（GenAI）語言模型的專門通函，重點警示持牌機構在使用GenAI撰寫研究報告或回覆客戶時，必須妥善管理AI的「幻覺風險」、資訊洩漏、算法偏見，以及違反知識產權等問題，特別強調在高風險用例中必須保持「人為監察」。

隨後在2025年，面對本地網絡安全事故大幅飆升，黑客廣泛利用AI進行深度偽造與精密系統漏洞攻擊，證監會隨即展開審查，並於2026年年中發出網絡安全通函，要求機構升級防禦架構以對抗AI驅動的網絡威脅。同年，證監會聯同金管局、保監局等推出「GenAI沙盒++」。

券商應用AI情況

智能客服	使用AI聊天機器人提供24/7即時賬戶查詢及市場數據解答
智能投顧	運用演算法根據客戶風險偏好，自動生成個人化資產配置建議
精準推送	分析客戶交易習慣，自動推薦相關的研究報告或股票資訊
報告生成	利用AI快速梳理財報與大數據，自動撰寫研究報告初稿及摘要
主動交易	測試代理式AI，在預設風控門檻內由系統自發捕捉信號並下單
遠端開戶	應用AI人臉識別與大數據比對，快速完成KYC（Know Your Customer／認識你的客戶）身份認證與反洗錢篩查
市場監察	實時監控交易盤路，自動偵測並攔截操縱市場的異常交易行為

AI針▶梁鳳儀表示，證監會的監管方針會聚焦於安全及負責任地使用方

